Mobile based Text Image Recognition using Deep Learning Approach

Saw Zay Maung Maung, Nyein Aye
sawzaymaungmaung@ucsy.edu.mm, *nyeinaye@gmail.com*

Abstract

Recognizing text image from mobile phone is a challenge task for limited capacity and processing power. And also the accuracy of the system is important for text image recognition system. In this system, we aimed to develop a Text Image Recognition System for mobile environment using Myanmar Character Dataset. Firstly, the image captured from mobile phone's camera and then segment each connected character using Connected Labeling Algorithm. After that the segmented characters input into the Convolutional Neural Network by passing layer by layer to get feature maps for recognizing the words in a given text image.

Keywords: *Connected Labeling Algorithm, Convolutional Neural Network*

1. Introduction

There are many image processing's utilities to be applied in recognizing characters in a text image. Earlier day, many Optical Character Systems is developed for computer system such as Laptop and Desktop computer environments and many applications are seen in real world environment using OCR. Nowadays, we use smart phone in every day. With smart phone, we can develop much smart system like the system of earlier day to take many advantages in advanced technology. The mobile computing devices include a camera so that software in the device can use this device to take pictures of the images such as a hand written text as well as printed text. For mobile devices, many image processing tasks can be applied to develop many smart applications using OCR.

With innovation technology, many modernized algorithms are applied in image classification and recognition system. Deep learning is one kind of machine learning's subfields and it gave the system with high accuracy. The Convolutional Neural Network is one kind of Deep Learning Algorithm and it applied in much image recognition system to get the results with higher accuracy. There are many deep learning frameworks (Tensorflow, Caffe, CNTK, PyTorch, Keras, and Deeplearning4j) to develop deep learning application. In this paper, we used the Tensorflow and Tensorflow Lite framework to use in mobile environment. Mobile phone includes many devices (Sound, Camera, and Internet) to develop many real time applications. By combining deep learning with mobile OS, mobile users can get many advantages with high accuracy and high availability.

The organization of this paper is as follows. Section II provides the related works. Section III shows the step by step image processing in android environment. Convolutional Neural Network is presented in Section IV. In Section V, we provide the system overview and detail explained of Deep Learning. Section VI provides the experimental results of this system and Section VII shows the conclusion of this paper.

2. Related Works

An Kohonen Neural Network based character recognition system explained and gave in **[1]**. This paper provided an framework for object oriented modeling and explained the challenges faced and the feature extraction method to detect characters. The OCR implementation is developed to learn Indian regional languages, as the number of characters including vowels, consonants and complicated letters are very much similar to most of the other Indian languages. In **[2]**, This paper aims to recognize and produce into an editable text from the image using Optical Character Recognition (OCR) method with Tesseract, an OCR engine which along with all image processing suite, is installed in the android app.

[3] proposes a method for Tamil Text detection in natural scene picture. The Maximally Stable Extreme Regions (MSERs) is extracted as character candidates using the strategy of minimizing regularized variations. By using a single link clustering algorithm, Character candidates are merged into text candidates where distance weights and clustering threshold are learned automatically by a novel self-training distance metric learning algorithm. The documents in this application scanned as images and once the image is scanned the data from the image

is extracted automatically and will be shown in the application as text. Then the text message is given to the translator tool which will convert
the Tamil text into English Text message.

The text region in document are scanned properly and then it segments the characters in **[4]**. After preprocessed and recognized a given text image, it will convert the English text into Marathi in translation process. Neural networks have been applied to various pattern classification and recognition. The input to a Kohonen algorithm is given to the neural network using the input neurons. And that input neurons get easily trained and having properties like topological ordering and good generalization. It uses smart mobile phones of android platform.

Myanmar text extraction and recognition from warning signboard images taken by a mobile phone camera is presented in **[8]**. The horizontal projection profile, vertical projection profile and bounding box are used to segment Myanmar Characters, The blocking based pixel count and eight-direction chain codes features are used in template matching method for recognition.

3. Image Processing

Image Processing is the task of analysis and manipulation of a digitized image to improve its quality. To use image processing in mobile phone, the image is firstly scan from real image objects using mobile phone camera. After scanning, the digitized image is preprocessed that is suitable for a particular application. In this system, we applied black and white conversion and morphology close processing to enhance the image quality. To segment each character or word, Connected Labeling Algorithm is used to get bounding box characters or words. The algorithm consists of two passes as follow:

On the first pass:

1. Scan column pixels first, then row pixels (Raster Scanning)
2. If the pixel is the foreground pixel
    a. Get the neighboring pixels of the current pixel
    b. If there are no neighbors, uniquely label the current pixels and continue
    c. Otherwise, find the neighbor with the smallest label and assign it to the current pixel
    d. Store the equivalence between neighboring labels

On the second pass:

1. Scan column pixels, then by row pixels
2. If the pixel is the foreground pixel
    a. Relabel the element with the lowest equivalent label

The method is applied in mobile phone and the result is shown in the figure 1.

Figure 1. Result of Using Connected Labeling Algorithm Method

4. Convolutional Neural Network

Convolutional neural networks convert the input data by passing through each layer of the network seamlessly to extract automatically the features of the images. In CNN, there are many different types of layers such as Convolution layer , Rectified Linear Unit , Pooling, Dropout , Fully connected layers.

Convolution layer in figure 2, also called a feature extractor, extracts features from the input image, tries to find them all places in the image by using a matrix called filter(Kernel).

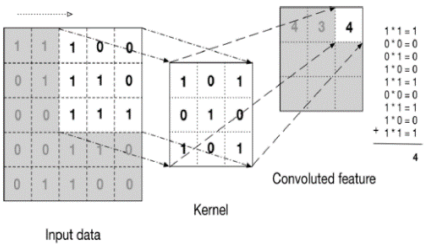


Figure 2. Convolution Layer

In figure 3, after each convolution, the Rectified Linear Unit (ReLU) is applied to produce an output by using an activation function of the neurons.

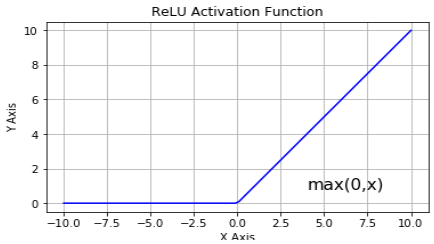


Figure 3. ReLU activation Layer

The Pooling layer, in figure 4, reduces the dimensionality of each image from the previous layers, by preserving the most important features.

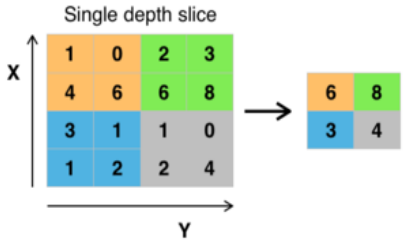


Figure 4. Pooling Layer

Dropout layers set their activation to zero to "drops out" a random set of neurons in that layer with the aim of reducing overfitting. In Full Connected Layer, every neuron is connected to every neuron by layer to layer.

The concept of convolutional neural networks is introduced in **[6]**, to find locally sensitive and orientation-selective nerve cells in the visual cortex **[7]**. A network structure is designed to extracts relevant features implicitly, by restricting the neural weights of one layer to a local receptive field in the previous layer to obtain feature map. By reducing the spatial resolution of the feature map, a certain degree of shift and distortion invariance is achieved **[7]**. Also, the number of parameters is significantly decreased by using the same weights for all features in the feature map **[5]**.

The following figure 5 is architecture of CNN for classifying character image

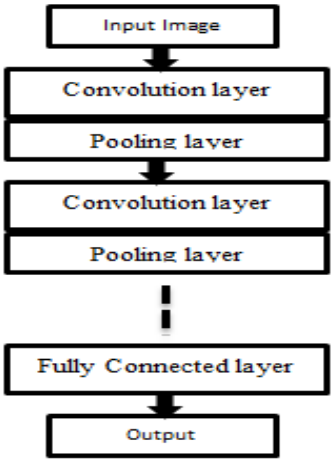


Figure 5. Convolutional Neural Network

4.1 Tensorflow and Tensorflow Lite

TensorFlow is an open source deep learning and machine learning software library for high performance numerical computation. It allows easy deployment of computation across a variety of platforms (CPUs, GPUs, TPUs), and from desktops to clusters of servers to mobile and edge devices.

TensorFlow Lite is TensorFlow's lightweight solution for mobile and embedded devices. It enables on-device machine learning inference with low latency and a small binary size. To apply deep leaning into mobile devices, we first need to convert ternsorflow trained model into tensorflow lite model (.tflite) by using tensorflow lite converter. The following figure 6 show the description of this process for TensorFlow Lite Architecture:

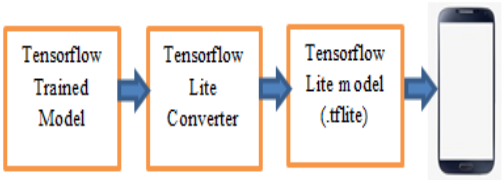


Figure 6. Processes for TensorFlow Lite

5. System Design and Description

There are two main processes in this system: **Training** and **Recognition**. Training is applied in Desktop Computer to produce a trained model that is transferred into Mobile Phone later. Recognition takes place in Mobile Phone using many image processing utilities. To process image captured from mobile camera, this system is firstly converted original image into grayscale image. And then apply Sobel edge detection algorithm to enhance character edges. The Sobel Operator computes an approximation of the gradient of an image intensity. The horizontal changes are computed by convolving image I with a kernel G_x with odd size. For example for a kernel size of 3, G_x would be computed as:

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I$$

The vertical changes is computed by convolving image I with a kernel G_y with odd size. For example for a kernel size of 3, G_y would be computed as:

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * I$$

An approximation of the gradient at each point of the image calculated using the following equation:

$$G = \sqrt{{G_x}^2 + {G_y}^2} \quad (1)$$

After that the thresholding operation is applied to convert binary image using the following formula:

$$dst(x,y) = \begin{cases} maxVal & \text{if } src(x,y) > thresh \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

And then apply Morphological Closing to remove small holes (dark regions) using the following equation:

$$A \bullet B = (A \oplus B) \ominus B \quad (3)$$

Closing is formed by first dilating a image A, after which this dilated set, A ⊕ B, is eroded.

Finally the Connected Labeling Algorithm is used to find the characters boundary. Now, the processed characters are ready for character recognition using Convolutional Neural Network. The Detail of Training and Recognition is described in subsection 5.1. The overview of this system design is shown in figure 7.

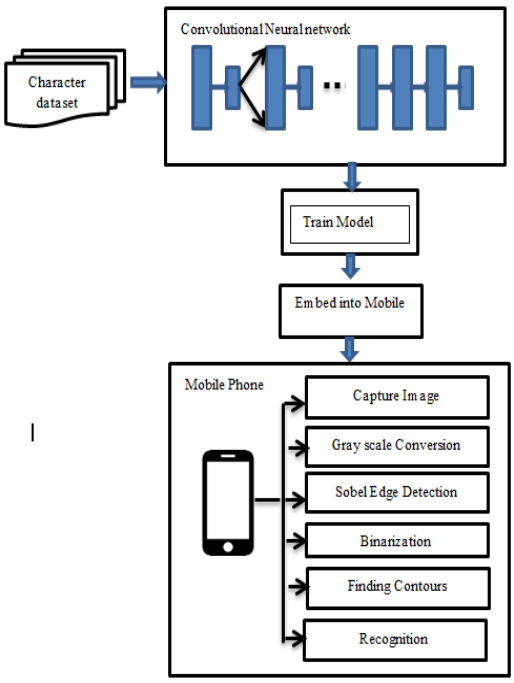


Figure 7. Overview of System Design

5.1 Training and Recognition

To train deep learning model for this system, we used Myanmar character image dataset. The Myanmar Character images Dataset used in this system are the book from Burmese Phrasebook (Survival Phrases in Myanmar) that is written by Naing Tinnyuntpu.

Example Survial Phrases are as follows:

The book include more than 300 distinct syllables words and use most important 250 syllables words to train this system. And then we generate the variations of each syllable by using Affine Transformation (rotation, scaling, zooming, shearing) to produce 12500 syllables words. After that we split the dataset as training data and test data. The following figure shows a design of Convolutional Neural Network for classifying Myanmar Character of this system:

where **CONV** – Convolution Layer,
ReLU – Rectified Linear Unit,
POOL – Pooling Layer,
FC – Fully Connected Layer,
SOFTMAX – Softmax Layer.

Layer Type	Output Size	Filter Size
Input Image	50x50x3	
CONV	50x50x20	5x5,K=20
ReLU	50x50x20	
POOL	25x25x20	2x2
CONV	25x25x50	5x5,K=50
ReLU	25x25x50	
POOL	12x12x50	2x2
FC	500	
ReLU	500	
FC	250	
SOFTMAX	250	

Figure 8. Design of Convolutional Neural Network

After the training is finished from Desktop environment, we get a tensorflow model (pb file) and then the tensorflow model is converted into tensorflow

lite model (tflite file) to embed into mobile environment for recognition.

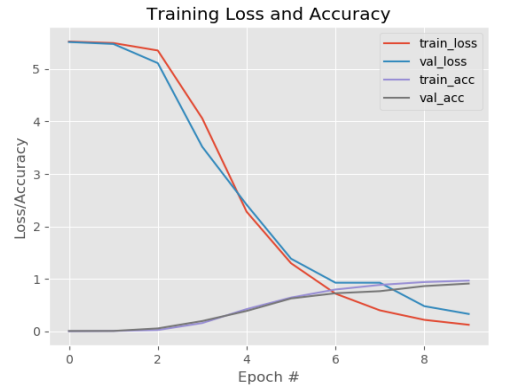
6. Experimental Result

As an experimental result, we test and train the survival words with iteration=10, iteration=20, iteration=50. The accuracy result is show in the following table 1.

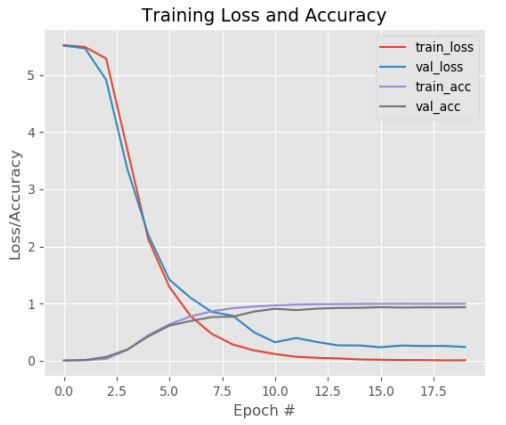
Table 1: Accuracy

Iteration	Precision	Recall
10	0.93	0.91
20	0.94	0.94
50	0.95	0.94

The following figure shows the loss and accuracy for iteration 10 and 20, respectively, during training process.



(a)



(b)

Figure 9. Loss And Accuracy
(a) iteration=10 (b) iteration=20.

7. Conclusion and Future Works

Text Image Recognition System for mobile environment is not an easy task because of limited capacity. Nowadays, the usages of deep learning are increased in the field of computer vision and image analysis because deep learning provides the results the higher accuracy than many traditional approaches. Mobile phone includes many devices (Sound, Camera, and Internet) to develop many real time applications. By combining deep learning with mobile OS, mobile users can get many advantages with high accuracy and high availability. In this paper, we used Convolution Neural Network to train and recognize Myanmar Character Images to get high accuracy. We get high availability every time and everywhere by transferring the trained model into mobile environment. In future direction, we can apply the recognized text into editable format for some OCR applications. And also provides translation into another language by recognizing the semantics of the combined images. Moreover, we can apply in the areas of image to speech application development.

References

[1] R.Shukla, "Object oriented framework modeling of a Kohonen network based character recognition system", Computer communication and informatics international conference (ICCCI), p 93-100, 2012.

[2] Ishita Pal, Mohammadraza Rajani, Anusha Poojary, Priyanka Prasad, "Implementation of Image to Text Conversion using Android App", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 6, Issue 4, April 2017.

[3] K. Jayasakthi Velmurugan, M.A. Dorairangaswamy,"TAMIL CHARACTER RECOGNITION USING ANDROID MOBILE PHONE", ARPN Journal of Engineering and Applied Sciences, VOL. 13, NO. 3, FEBRUARY 2018.

[4] Mayuri B Gosavi, Ishwari V Pund, Harshada V Jadhav, Sneha R Gedam," Mobile Application with Optical Character Recognition Using Neural Network", International Journal of Computer Science and Mobile Computing, Vol.4 Issue.1, January- 2015.

[5] Lecun Y, Bottou L, Bengio Y, Haffner P, "Gradient-Based Learning Applied to Document Recognition", in Proceedings of the IEEE, 1998.

[6] Lecun Y, Bengio Y, "Convolutional Networks for Images, Speech, and Time-Series in The Handbook of Brain Theory and Neural Networks", MIT Press 1995.

[7] HAYKIN S,"Neural Networks: A Comprehensive Foundation", second edition. Prentice Hall 1999. Chapter 4 Multilayer Perceptrons, pp. 156 − 255.

[8] Kyi Pyar Zaw, Zin Mar Kyu, "Camera Captured based Myanmar Character Recognition Using Dynamic Blocking and Chain Code Normalization", International Journal of Scientific and Research Publications, Volume 8, Issue 8, August 2018